

Issued Patent

[51]Int.Cl⁶

G06F 3/023

[19]中华人民共和国专利局



[12] 发明专利说明书

[21] ZL 专利号 94110164.9

[11] 授权公告号 CN 1036548C

[45]授权公告日 1997年11月26日

[22]申请日 94.4.1 [24]颁证日 97.9.20

[21]申请号 94110164.9

[73]专利权人 张金岭

地址 110001辽宁省沈阳市和平区中山路50号
和平区通宝文化用品经营部

[72]发明人 张金岭

[74]专利代理机构 沈阳市第二专利事务所

代理人 于飞

[56]参考文献

CN921101716 1993. 7. 13 G06F3 / 023

CN921140517 1993. 6. 9 G06F3 / 023

审查员 张静海

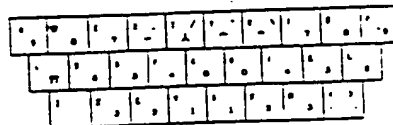
权利要求书 2 页 说明书 11 页 附图页数 2 页

[54]发明名称 一种用数字将汉字输入计算机的方法

[57]摘要

本发明涉及一种计算机的汉字输入方法，是一种新颖的计算机输入方法，它是根据汉字的结构转化为阿拉伯数字运用计算机标准键盘输入的方法。本发明的特点是在计算机标准键盘上设定二套 0—9 数字键，将群体汉字结构分为上下结构，左右结构，四面包围、半包围、独体结构五类，再按结构将汉字的字首笔划数细分为 10 个区，将汉字转化为数字，进行单字和词组的输入。本发明简单易学、应用广泛。

数字键盘分布图(续3)



权 利 要 求 书

1.一种用数字代码将汉字输入计算机的方法，其特征在于：

A、利用计算机标准键盘，在标准键盘上设定二套 0—9 数字键，左手数字键 1、2、3、4、5、6、7、8、9 对应原键位 V、C、X、F、D、S、E、W、Q，右手数字键 1、2、3、4、5、6、7、8、9，对应原键位 B、N、M、J、K、L、I、O、P，首部代码“一”，“人”，“亠”，“艹”“ ”对应原键位 R、T、Y、U、A，在 Q—M 键盘中，TGV 键以左的键帽为左手键，YHB 键以右的键帽为右手键，其中 RTYU 键为特定 O 键，R 键代表“一”，包括“亠”、“ ”、“厂”，“ ”、T 键代表“人”字头，包括“八”，Y 键代表“亠”，包括“一”，U 键代表“艹”，在输入过程中，遇 0 区汉字符合上述条件字首优先敲击特定 RTYU 键，A 键为特定 1，A 键代表“ ”、“心”、“灬”，Z 键为技术处理键；

B、汉字取码方法如下：

(1) 将群体汉字结构分类五类：上下结构、左右结构、四面包围、半包围、独体结构；

(2) 按结构将汉字的字首的笔划数细分为 10 个区域：上下结构的汉字分为两个区，字首在三笔划以内的划为 0 区，字首在四笔划以上的划为 1 区，左右结构的汉字划为六个区，即左面两笔划为 2 区，左面三笔划为 3 区，左面四笔划为 4 区，左面五笔划以上为 5 区，“口”旁和“𠔁”同四面包围汉字统一组成第 6 区，第七区由半包围结构组成，8 区是由“彳”、“扌”组成，9 区是由不能进行上述分类的独体字组成。凡是落入该数字区域的汉字部首就是该汉字的第一数字代码，余下部分汉字在按结构分类编码输入；

(3) 字尾编码输入方法如下：字的末笔划统分横竖撇捺折五种，

笔划无交叉用 1—5 代码输入，笔划有交叉用 6—0 代码输入，“口”字为特定末笔，代码为 0；

(4) 根据被编码字首笔划，字首无交叉笔划归左手数字键，字首有交叉笔划归右手数字键，凡是被编码字首在字的右侧一律归右手数字键，凡是被编码由左上半包围和三面包围无交叉字首归左手数字键，其它半包围形式归右手数字键；

利用上述键位代码输入步骤如下：

(1) 单字输入的方法是，根据汉字结构确定第一码，再依次根据结构确定第二码，第三码，末笔码为第四码，被编码字首无交叉笔划归左手键，有交叉笔划的归右手键；

(2) 独体结构汉字的输入方法是，独体结构汉字不进行结构分类，第一码特定为 9，依次根据结构和末笔确定第二码，第三码，第四码；

(3) 词组输入方法：两字词采用汉字编码每字取头两码编码，三字词采用第一，第二字取头一码，第三字取头两码，四字词每字取头一码，多字词取头三字和最后一字的第一码，词组输入时，所有被编码字首有交叉笔划归右手键，字首无交叉笔划归左手键。

字键，

X、

、9、

人”，

言中，

其中

”、

代表

特定

末处

面包

结构

划以

区，

区，

结构

本字

：码，

种，

说明书

一种用数字将汉字输入计算机的方法

本发明涉及一种计算机的汉字输入方法，是一种新颖的计算机输入方法，它是根据汉字的结构转化为阿拉伯数字运用计算机标准键盘输入汉字。

目前，计算机汉字输入方法种类繁多，按输入方式，一般分为音码、形码、音形结构码、声码、声控输入、手写输入等形式。纵观这些编码方法普遍存在规则复杂，难学难记的弊端。在诸多输入方法中用数码输入是较理想的输入方式，有代表性的是区位码和电位码，但这两种方法数字与汉字之间无规律可循。本发明就是研究汉字与数字之间内在联系，寻求数字变化同汉字结构演变的规律，将汉字用数字描绘出来输入计算机。

本发明的特点是学习容易，操作简单。汉字分类小学生用几小时就可掌握，成年人几个小时即可学会，操作对象只是0—9数字键，简单易学。本发明应用范围广，不仅是理想的电子计算机汉字输入方法，而且可以应用于信息检索、档案管理、图书管理、字典查字、机器翻译有其他各类需要编码的领域。有利于计算机应用的推广和普及。

本发明通过如下技术方案实现：通过汉字的结构转化为数字再利用计算机标准键盘输入计算机。

汉字转化为数字的方法，依据具体汉字结构的差异和汉字内

部结构变化规律进行数字编码，配合键盘，具体方法是：

(1)、将群体汉字按结构分为五类：上下结构、左右结构、四面包围、半包围、独体结构。

(2)、按结构将汉字的字首的笔划数细分为10个区域：上下结构的汉字分为两个区，字首在三笔划以内的划为0区，字首在四笔划以上的划为1区，左右结构的汉字划为六个区，即左面两笔划为2区，左面三笔划为3区，左面四笔划为4区，左面五笔划以上为5区，“口”旁和“彡”同四面包围汉字统一组成第6区，第7区由半包围结构组成，8区是由“扌”，“扌”组成，9区是由不能进行上述分类的独体字组成。凡是落入该区域的汉字就是该汉字的第一编码，余下部分汉字再按结构分类编码，以此类推（见编码表）。

(3)、通过若干次分类编码汉字就剩余字尾，字尾编码，方法如下：字的末笔划统分横竖撇捺折五种，用1—5表示，有交叉的末部首横竖撇捺折用6—0表示，“口”字为特定末笔，编码为0。

(4)、末笔码可编第二码，第三码和第四码。

(5)、汉字转化为数字，输入单个字，不仅为四个数码，也可以一数码输入，二数码输入，三数码输入。

(6)、汉字转化为数字，不仅可以单字输入，还可以两字词输入，三字词输入及多字词输入。

(7)、利用计算机标准键盘，在标准键盘上设定二套0—9数字键，在Q—M键盘中，IGV键以左的键帽为左手键，ZHB键以右的键帽为右手键，其中RTYU键为特定0键，R键代表“一”，包括“一”。

“一”、“丿”、“㇏”，I键代表“人”字头，包括“八”，Y键代表“宀”，包括“冖”，U键代表“艹”，在输入过程中，遇0区汉字符合上述条件字首优先敲击特字RTYU键，A键为特定1，A键代表“木”、“心”、“灬”，Z键为技术处理键。

(8)、根据被编码字首笔划，字首无交叉笔划归左手数字键，字首有交叉笔划归右手键，凡是被编码字首在字的右侧一律归右手数字键，凡是被编码由左上半包围和三面包围无交叉字首归左手数字键，其它半包围形式归右手数字键，单字输入时，第四码跟随第三码手位。

(9)、词组输入方法：两字词采用汉字编码每字取头两码编码，三字词采用第一、第二字取头一码，第三字取头两码，四字词每字取头一码，多字词取头三字和最后一字的第一码。

10、词组输入时，所有被编码字首有交叉笔划归右手键，字首无交叉笔划归左手键。

以下通过汉字编码实例和键盘应用进一步说明本发明的具体操作方法。

结合汉字结构编码表一，末笔编码表二，键盘分布表三，详细说明本发明。

一种用数字将汉字输入计算机的方法，将汉字的类型设定归纳为5种类型划分10个区域，5种结构类型，1、上下结构，2、左右结构，3、全包围结构，4、半包围结构，5、独体结构。又将汉字分成5个类别划分10个区，其中上下结构分成2个区即0区和1区，0区指上下结构，上部笔划在三划以内的(包括3划)，1区指上部笔划4划以上的(包括4划)的汉字，左右结构划分为5个区。

由左部2笔划组成的汉字定为第二区，由左部3笔划组成的汉字定为第三区，由左部4笔划组成的汉字定为第四区，由左部5笔划以上(包括5笔划)组成的汉字定为第五区，由左部“扌”和“扌”单独拿出来组成左右结构的第八区，第六区是由四面包围形的汉字来组成，由于四面包围的汉字比较少，所以将左右结构汉字“口”字旁的也归为第六区，“彳”归为第六区一是考虑五区的汉字分配的比较多，二是为了左右手键盘分配的合理特定的，第七区是由半包围型汉字组成的，它不考虑外部包围形体，只要满足二面及三面包围的即可，第九区是不能进行上述分类的不规则的汉字(见表一)。

只了解字的结构进行汉字编码是不完全的，因为每一个汉字都有字尾也称末部首，例“寄”字的字尾是“口”字，每一个字的字尾是不同的，这就需要对字尾也要进行编码，在进行笔划编码时，只考虑笔划的运笔方向，而不计较轻重长短，根据其使用频率，便于记忆和应用的原则，将笔划分为横、竖、撇、捺、折5种，又因每一字的末笔存在于交叉与非交叉偏旁部首之中，为了增加编码范围将横、竖、撇、捺、折扩展为有交叉的横、竖、撇、捺、折和无交叉的横、竖、撇、捺、折，因此将5种末笔划码增加到10种(见表二)，确定末笔时有两种情况需要注意，一是“乚”、“戈”这类字末笔为撇，二是某些字被编码切断时确定字尾交叉与否，如“虫”字作为字尾出现末笔为9，如以单字出现末笔为4(因被编码切断)，另外有一部分字的编码信息量不足这时再取字尾首笔笔形码补充编码，字尾首笔码编法是编码本笔无交叉用1—5，本笔有交叉6—0表示。

在编码过程中，首先是查看汉字的基本结构，确定汉字的类型，然后查看部首的笔划数量，按分类编码表的要求即可确定第一码，第一码确定后再查看余下部分汉字结构，这时会出现两种情况，一是还能继续进行上述结构分类化区，如遇这种情况则应继续按编码分类方法分类，二是遇到土、女、木、小、巾、口、人等汉字字尾时应按末笔编码表确定末笔码，如汉字字尾的无交叉笔划，末笔则按“一、丨、丿、丶、乙”用数字1、2、3、4、5来代表它，如汉字字尾有交叉笔划则用数字6、7、8、9、0来描绘，例如，某汉字字尾是“小”字无交叉笔划，末笔是“丶”按为4，这时末笔编码为4，又如，某汉字的字尾是“土”土字有交叉笔划，末笔是横但有交叉笔划则横为6，这个末笔码为6，末笔码可编第二码、第三码和第四码，不能编第一码，第三码的确定有三种情况，如该汉字经两次结构分类划区后还能进行分类划区，则继续分类划区编码，自然状态编码应为3码或4码，如编完第三码、第四码就按末笔划码法编码。

例，如汉字的“字”，先看“字”字由上下结构组成，上部的“宀”三笔划划分为0区，第一编码为“0”，再编第二码，看余下部分汉字“子”字的末笔为横，字尾有交叉笔，查表得知末笔码为6，即第二码为6，第三码为字尾首笔码，字尾首笔顺为“折”编码为5（字尾首笔码本笔无交叉编码1—5，本笔有交叉编码6—0），“字”字的编码为065。

例，“例”字去掉字首“亻”后是“奇”字，“奇”字去掉字首“大”字是“可”字，“可”字去掉“丁”是“口”字，“口”字为特定字尾编码为0，因此“奇”字的编码为0070。

的类
定第
两种
则应
口、
无交
4、
描绘、
34、
笔划、
第
中情
续
第

例：“唱”字为左右结构，左部为“口”，编码为6，余下部分是“昌”字，“昌”字为上下结构，上部“日”笔划是四笔以上，编码为1，余下“日”字，“日”字按结构分为第6区，则第三码定为6，“日”字去掉外框之后，仅剩一笔“一”，按字尾编码可知，编码为1，唱字的编码6161。例：“编”字，按结构分类为左右结构，左边“纟”三笔划属于第三区汉字，第一编码为3，余下部分是“扁”字，“扁”字为半包围结构划为第七区，第二编码为7，“冂”部首同样是半包围结构划为第七区，第三编码为7，最后剩余“艹”末笔为竖有交叉笔编码为7，因此“编”字的编码为3777。以上几例可以看出：“寄”字与“唱”虽然都有“口”，但编码数字却不同，“寄”与“奇”的字首不同，但编码相同，这说明一个实质问题就是：要掌握字的结构和字首复杂程度的来进行编码。

计算机键盘键位分配

二部
看
末
为
编
三
字、

由汉字的结构进行数字编码是根据近似模糊数学的原理，仅对字的结构进行分类，而不对字的每一笔顺细分化，将造成重码率较高，这只能应用在对输入或查找汉字要求速度不高的地方使用。为了减少重码率增加输入速度我们对标准键盘0—M输入键进行以下利用(见表三)，设计两套0—9数字在键盘上，在0—M键盘中，TGV以左0—9为左手用键，用以操作被编码无交叉部首，YHB以右为右手用键，用以操作被编码有交叉部首，其中RTYU键为特定0键，以分散0区编码密度，A键为特定1键，以分散1区编码密度，Z键为技术处理键，由于两套0—9数字可以交替操作，编码容量可达16万，大大地减少了重码率。本编码法有很多汉字编码

虽然相同，但通过两套数字交替操作都能做到不重码输入。

汉字的编码输入：

一、上下结构汉字的编码及输入

上下结构的汉字在园标一级字库中占25%。其中字首在三笔划以下的(简称0区以下同)占12%。由于在群体汉字中进行一次分类反映到上下结构占70%，而且字首在反映到0区的占50%，因此对0区进行细分化尤为重要。在0区除有两个0以外，还分别设立RTYU四个键来承担0区的任务，R键代表“一”包括“一”、“乚”、“厂”、“乚”，T键代表“人”字头包括“八”，Y键代表“宀”，U键代表“艹”。在输入过程中如遇0区汉字优先击符合上述条件字首的键位，而不去击0键。0区汉字编码，例：“分”字，第一码0(结构码)，第二码3(末笔码)，第三码5(字尾首笔码)，“分”字编码035。例“京”字，第一码0(结构码)，第二码0(结构码)，第三码4(末笔码)，“京”字编码004。例：“蒋”字，第一码0(上下结构)，第二码3(左右结构)，第三码0(上下结构)，第四码9(末笔码)，“蒋”字编码0309。两套数字键分配，被编码字首无交叉笔划归左手数字键，有交叉部首归右手数字键。

上一结构字首四笔划以上(简称1区以下同)占13%，该区汉字字数较多字形也比较复杂。因此我们将“A”键也作为特定1键，把“竹”头“心”“灬”底规结到A键而不使用1键，其它1区汉字按规则要求输入。该区编码又略有差别，对于上部为整体结构下部属于天常规律变化的汉字同0区的编码方法一样，如遇上部左右两部分组成的字按下列方式定码，第一区的汉字第一码为1，第二码则查看上部按左右结构编码法编码，第三码以后则继续按

规则编码。1区汉字编码，例“爸”字，第一码1(结构码)，第二码5(末笔码)，第三码5(字尾首笔码)，“爸”字编码155。例“型”字，第一码1(结构码)，第二码2(结构码)，第三码6(末笔码)，“型”字编码126。例“想”字，第一码1(上下结构)，第二码4(左右结构)，第三码6(全包围结构)，第四码1(末笔码)，“想”字编码1461。两套数字键分配，被编码字首无交叉笔划归左手数字键，有交叉部首归右手数字键。

二、左右结构汉字的输入

左右结构的汉字占60%，数量最多，因此划分区域也最多，共化为2、3、4、5、6、8六个区。其中左面两划以下及右面“刂”“冫”划为2区，左面三笔划为3区，左面四笔划及右面“攵”划为4区，左面五笔划以上及右面“页”划为5区，“口”“彡”划为第6区，“彳”“亍”划为第8区。2区汉字编码，例“他”字，第一码2(结构码)，第二码0(末笔码)，第三码0(字尾首笔码)，“他”字编码200。例“到”字，第一码2(结构码)，第二码0(结构码)，第三码6(末笔码)，“到”字编码206。例“倒”字，第一码2(左右结构)，第二码2(左右结构)，第三码0(上下结构)，第四码6(末笔码)，“倒”字编码2206。例“摄”字，第一码8(左右结构)，第二码1(上下结构)，第三码2(左右结构)，第四码9(末笔码)，“摄”字编码8129。左右两套数字键位分配：凡是被编码的字首无交叉笔划的归左手数字键，凡是有交叉笔划的归右手数字键，凡是被编码字首在字的右侧一律归右手数字键，特定2区“亻”为右手数字键。

三、四面包围结构汉字的输入

四面包围结构的汉字在群体汉字中数量较少，仅占1%。但在个体汉字结构分类中又有一定量的比重，因此划为6区。我们将“口”“与”也划为6区以补偿该区群体汉字的不足。6区汉字编码：例“园”字，第一码6(结构码)，第二码6(末笔码)，第三码1(字尾首笔码)，‘园’字编码661。例“哈”字，第一码6(结构码)，第二码0(结构码)，第三码0(末笔码)，‘哈’字编码600。例“锯”字，第一码6(左右结构)，第二码7(半包围结构)，第三码0(上下结构)，第四码0(末笔码)，‘锯’字编码6700。两套数字键位分配：凡是被编码的字首无交叉笔划的归左手数字键，凡是有交叉笔划的归右手数字键。

四、半包围结构汉字的输入

半包围结构汉字占4.7%，凡是满足两面、三面包围的汉字均属7区，不考虑包围体结构。7区汉字编码：例“辽”字，第一码7(结构码)，第二码2(末笔码)，第三码5(字尾首笔码)，‘辽’字编码725。例“远”字，第一码7(结构码)，第二码0(结构码)，第三码5(末笔码)，‘远’字编码705。例“痛”字，第一码7(半包围结构)，第二码0(上下结构)，第三码7(半包围结构)，第四码7(末笔码)，‘痛’字编码7077。两套数字键的分配：凡是被编码由左上半包围和三面半包围无交叉字首归左手数字键，其余半包围形式归右手数字键。

五、独体结构汉字输入

独体结构汉字是指不能进行上述分类的汉字。对这些汉字在编码时第一码设定为9即可，如“也”字，第一码为0，第二码为0，第三码为0，“也”字的编码900。例“主”字，第一码为9，

31%。但
区。我们
，6区汉字
3)，第三
一码6(结
字编码600。
内)，第三
1，两套数
字键，凡

围的汉字
字，第一
3)，‘辽’
结构码)，
一码7(半
)，第四
凡是被编
，其余半

些汉字左
第二码为
一码为9，

第二码为6(末笔码)，第三码4(字尾首笔码)，‘主’字编码964。
另外一些汉字从上到下或从左到右有一笔相连，对这些汉字第一
码也编码为9，其余看整体结构进行拆分。例，‘我’字只有一横
相连化分为准左右结构，编码948。例‘里’字只有一竖相连化
分为准上下结构，编码916。对于独体字上部有两点也划为本区
例，‘单’字编码为9017。

两字词的编码方法：两字词在编码时取组成两字词单字的前
两码组成。例，‘编码’一词，‘编’字前两码是3、7(3左右结
构，7半包围结构)，‘码’字前两码是5、1(5左右结构，第五区，
1末笔码)，则“编码”一词的编码是3751。击键方式：全为左手。

三字词的编码方法：三字词在编码时取组成三字词的第一、
二单字的第一码，取最后一字的前两码。例，“辽宁省”一词，
‘辽’字为半包围结构，第一码为7，‘宁’字为上下结构，上
部是三笔划，第一码是0，‘省’字上下结构，上部是三笔画，
第一码为0，‘目’字为全包围结构，第二码为6，因此“辽宁省”
一词的编码是7006。击键方式：第一码为右手，其余左手。

多字词的编码方法：多字词在编码时是取前三个字的头一码，
取最后一字的头一码组成。例，“中华人民共和国”一词，‘中’
字第一码是6(全包围结构)，‘华’字上下结构，上部四笔划第
一编码为1，‘人’字为独体结构，第一码为9，‘国’字第一码
为6(全包围结构)，因此“中华人民共和国”一词的编码为6196。
击键方式：第一码右手，第二码右手，第三码左手，第四码为三
手。

本发明打破常规的汉字编码方法，是一项具有开拓意义的发明，它不仅适应于电子计算机的汉字输入，而且可用于其它各个需要汉字编码的领域。

5意义的发
三其它各个

汉字结构编码表

表一

编码	结构	编码根据	部首举例	例字	说明	部首分配
0	上下	上部三划以内 (包括三划)	一 一	字 065 寄 0070	查最上部 部首笔划	左手: 无交叉部首 右手: 有交叉部首
1	上下	上部四划以上 (包括四划)	木 日	查 111 常 107	查最上部部首笔划 及下部“心”“一”	同上
2	左右	左部两笔划右部 包括“刂 卩”	亻 彳 卩 卩	侧 2700 邻 204	右部“卩”“刂”	左手: 无交叉部首 “刂 卩”在右边归右手 右手: 特定“亻”
3	左右	左部三笔划“彳” “口”除外	土 彳 彳 彳	堤 3014 途 3052	“口”归6; “彳” 归3; “彳”归3	左手: 无交叉部首 右手: 有交叉部首
4	左右	左部四划 “彳”归4	月 日 三 彳	籍 4161 村 496	“彳”为5划, 但编 码为4 “文”为4	同 上
5	左右	左部五笔划以 上(包括五笔划)	血 角 示 示	螺 5104 称 504	左部5画及5画以上 编码为5, “页”为5	同 上
6	左右 全包	部首“口 彳” 四面包围	口 彳	喧 6001	左部为“口” 右部为“彳”	同 上
7	半包	二、三面包围	厂 辶	局 770 阔 728	满足二、三面包围 可不考虑外杠结构	左手: 特定左上及三面 无交叉笔半包围 右手: 其它包围为右
8	左右	左部“彳” “彳”	彳 彳	济 812	左部为“彳”的字 左部为“彳”的字	左“彳”右“彳”
9	独体		不 刀 火	夕 943 刀 935	不能进行上述结构的 独体字	

末笔编码表(表二)

笔	名称	横	竖	撇	捺	折
项	笔形	一	丨	丿	㇏	乙
码	部首无交叉	1	2	3	4	5
	部首有交叉	6	7	8	9	0
“口”字为特定末笔，编码为“0”						

数字键盘分布表(表三)

0 9	W S	E 7	R —	T 人	Y —	U —	I 7	O 3	P 9
A 竹	S 6	D 5	F 4	G O	H O	J 4	K 5	L 5	
Z	X 3	C 2	V 1	B 1	N 2	M 3	< .	> .	

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☒ BLACK BORDERS

☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☒ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☐ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)